

DOI:10.11918/202012047

考虑天气因素的给水管道漏损预测模型

侯本伟,肖恒圣,吴 珊

(北京工业大学 城市建设学部,北京 100124)

摘 要: 天气变化引起土壤含水率变化和变形等,导致气温变化与市政给水管道的破损事件存在相关性。为此,基于北方某城市给水管网破损事件数据和气温记录,分析不同的天气因素量化指标与管道破损事件的相关性。采用误差反向传播神经网络(BPNN)和基因表达式编程(GEP)方法,建立考虑天气因素的给水管管道漏损预测模型。根据案例城市过去 11 年的市政给水管网破损记录数据库、管道地理信息数据库和同期气温记录,分析 6 个天气因素指标(平均温度、冰冻指标、最大上升值、最大下降值、最大上升率、最大下降率)的内在相关性及其与管道破损事件的相关性;采用 BPNN 和 GEP 建立管道破损数(因变量)与 4 个自变量(代表性天气因素指标、管径、管龄、管长)的隐式和显式函数关系。应用隐式和显式两种模型预测案例城市给水管网未来 1 年的破损数,未考虑天气因素的模型预测决定系数分别为 0.65 和 0.60,考虑天气因素模型预测结果的决定系数分别为 0.78 和 0.88,预测精度提升率分别为 13% 和 28%。建立考虑天气因素的给水管管道漏损预测模型是合理有效的。

关键词: 给水管道;漏损预测;天气因素;冰冻指标;基因表达式编程

中图分类号: TU991 **文献标志码:** A **文章编号:** 0367-6234(2022)02-0008-09

Failure prediction model of water distribution pipelines considering weather factors

HOU Benwei, XIAO Hengsheng, WU Shan

(Faculty of Architecture, Civil and Transportation Engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: Due to the changes in soil moisture content and deformation caused by weather changes, there is a correlation between weather temperature and failures of urban water distribution pipelines. Based on the failure records of pipelines and weather temperatures of a northern China city, the correlations between different weather temperature indicators and water mains failures were analyzed. The failure prediction models of water distribution pipelines considering weather temperature factors were established using error back propagation neural network (BPNN) and genetic expression programming (GEP) methods. According to the failure records of water distribution pipelines, pipeline geographic information, and weather temperature records of the case city in the past 11 years, the correlations among pipeline failures and six weather factors were analyzed, including average temperature, freezing indicator, maximum increase, maximum decrease, maximum increase rate, and maximum decrease rate. BPNN and GEP were used to establish the implicit and explicit relationships between the number of pipeline failures (i. e., the dependent variable) and four explanatory variables (the selected weather temperature indicator, the diameter, age, and length of pipelines). The explicit and implicit models were used to predict the number of pipeline failures in the case city in the next year. The determination coefficients of the prediction results of the model without considering weather factors were 0.65 and 0.60 respectively, while those considering weather factors were 0.78 and 0.88, where the prediction accuracy increased by 13% and 28%. Therefore, it is reasonable and effective to establish a failure prediction model of water distribution pipelines by considering weather factors.

Keywords: water distribution pipeline; failure prediction; weather factor; freezing indicator; genetic expression programming

给水管管道漏损预测模型是给水管网运行维护管

理的重要研究内容,本研究的给水管管道漏损预测模型是通过分析、统计和学习管道漏失破损事件记录数据,建立管道漏损事件与管道自身因素、环境因素和管网运行因素等参数的显式/隐式函数关系,根据这种函数关系预测管道的未来破损数。基于给水管管道漏损预测模型可以制订管道更新改造,可为管道

收稿日期: 2020-12-15
基金项目: 国家重点研发计划项目(2018YFC1504602);
北京市教委科技计划项目(KM202010005031)
作者简介: 侯本伟(1984—),男,博士,副教授
通信作者: 吴 珊, wushan@bjut.edu.cn

漏损事件发生前的干预工作提供技术支撑,也可作为漏损事件发生后的漏失巡检工作提供依据。

漏损预测模型可大致分为3类,即物理模型、统计模型、数据挖掘模型。物理模型旨在发现管道破损现象下的物理机理,多是从材料力学角度出发,计算管道内外环境与管道之间的相互作用,从而判断管道在一定条件下是否发生漏损。该方法的分析过程需要严格精细的控制条件,主要用于漏损成本较高的重要输水管道或管道试验^[1-4]。统计模型通过历史数据分析,将管道破裂模式与管道属性参数、环境变量、管网运行参数联系起来^[2,5-7]。赵洪宾等^[8]根据实际给水管网的漏损数据,建立了给水管道的漏损预测的线性指数平滑模型和二次指数平滑模型;张宏伟等^[9]应用多元线性回归分析理论,建立给水管道的漏损发生时间为因变量,以管道接口类型、温度、埋深、直径、施工质量、运行压力和管道腐蚀指标为自变量的漏损预测模型,证明多元线性回归方法预测的平均相对误差为21%。数据挖掘模型利用人工智能和机器学习算法建立管道漏损的预测模型,是目前常用的给水管建模方法^[5,10]。根据模型表达形式,又可分为显式模型和隐式模型。Sattar等^[1]采用基因表达式编程(gene expression programming, GEP)建立管长、管径、管道保护方式、以前破损数与管道下次破裂发生时间的显式函数关系,误差分析结果表明,GEP模型预测结果稍好于误差反向传播神经网络(back propagation neural network, BPNN)模型。徐强等^[4,11-12]利用遗传编程和进化多项式建立了管道总漏损次数与管龄、管径、管长之间的函数关系,结果表明,两种模型预测性能相差不大。阎立华等^[13]采用BPNN构建了管道安全使用期与管径、压力、埋深、铺设年代、所在地区类型函数关系的预测模型,模型预测结果的相对误差均小于10%。

在相关研究中,管径、管龄、管长被认为是影响管道漏损的重要因素。除这些管道自身影响因素外,天气变化也是导致管道发生漏损的重要因素。Barton等^[14]对研究天气变化对管道影响的文章结论进行了综述,认为天气变化对管道漏损存在明显影响,而且铸铁管受天气影响极为明显,具体表现为夏季高温导致土壤水分蒸发致其收缩,然后降雨导致土壤膨胀从而造成管道发生漏损,冬季的极端低温导致土壤水分结冰致其膨胀,挤压管道导致漏损。邓晓婷^[15]统计了4个季度的漏损点数,发现一、四季度漏损点数比二、三季度增加16.6%。Rajani等^[16]通过定义16种不同的水温和气温变量指标,采用非齐次泊松分布模型分析不同分析步长(5~

90 d)内的气温变量指标与管道破损次数之间的关系,结果表明,最佳分析步长为30 d,对于不同管材需采用不同的气温变量进行分析。Kakoudakis等^[17]采用K-means聚类与进化回归多项式结合的算法分别建立了管道破损数与4个自变量(管长、管径、管龄、冰冻指标)、管道破损数与冰冻指标的显式函数关系,结果表明,考虑管道物理属性(管长、管径、管龄)之后的模型预测精度稍有提高。

通过上述文献可知,对于气温变化显著的城市,应建立考虑天气因素的预测模型研究天气因素对管道漏损的影响。由于不同城市气温变化和管网特征的差异,上述研究结论是否适用于中国城市给水管网的漏损预测,仍有待进一步研究。为此,基于中国北方某案例城市的数据,研究了天气因素量化指标与管道漏失事件的相关性,天气影响因素对模型预测精度提升的适用性,显、隐式函数关系构建模型的有效性。

1 研究区域介绍

研究对象是北方某城市给水管网干线,给水管网破损数据库记录了2008—2019年管道破损点的管材、管径、抢修类型及发生时间等属性;给水管网地理信息系统数据库记录了给水管道的管材、管径、铺设年份、管长等属性;气温数据库记录了每天的平均气温。案例城市给水管网干线主要由普通铸铁管(CIP)、球墨铸铁管(DIP)、钢管(SP)组成,表1为该市给水主干管道的基本统计信息。可以看出,普通铸铁管的单位管长破损数最高,破损点的抢修类型分为明漏、暗漏、工程漏,由于工程漏是由管道周边施工导致管道出现破损,与天气变化并无关系,故本研究中主要针对普通铸铁管的明/暗漏破损点建立漏损预测模型。

表1 给水干线基本情况
Tab.1 Characteristics of water mains

主干管道	管径/ mm	管长/ km	铺设 年份	明/暗漏破 损次数	单位长度破损 数/(次·km ⁻¹)
CIP	75~600	2 175	1920—1999	5 380	2.47
DIP	75~1 200	4 115	1990—2019	2 526	0.61
SP	75~2 200	572	1957—2019	246	0.43

对于普通铸铁管道,单位长度破损数随管径变化趋势见图1,破损次数随管龄变化见图2,每月破损数占比及平均气温随月份变化见图3,破损次数及管长随年份变化见图4,破损点每月平均埋深如图5所示。由图1可知,随着管径减小,单位长度破

损数呈现整体降低的趋势。图 2 的结果表明,管道漏损次数随管龄变化呈现先上升后下降的趋势,在管龄 20 a 时管道破损数最大,在管龄 65 a 后趋于稳定。图 3 的结果表明,普通铸铁管漏损次数与寒冷天气存在明显相关性,漏损次数随着气温降低而明显上升,原因可能是秋冬交替期间气温的骤降易导致管道破损数迅速上升,然后持续的低温会进一步导致破损数上升。图 4 的统计结果表明,在役普通铸铁管的总长度和年破损数呈现整体下降趋势。图 5 的结果表明,在破损数较多的 11 月至次年 1 月,这些月份的破损点平均埋深并没有明显小于其他月份。冰冻天气对管道的影响主要体现在土壤水分结冰致其膨胀,埋设在冻土层中的管道结构会受到约束导致漏损。实际中,管道上方土壤冻胀也可能影响位于冰冻线以下的管道,但目前尚缺乏土壤冻胀影响的物理机制、影响深度、管道结构内力和变形的增长率等相关研究。因此,并未根据冰冻线划分管道分组。

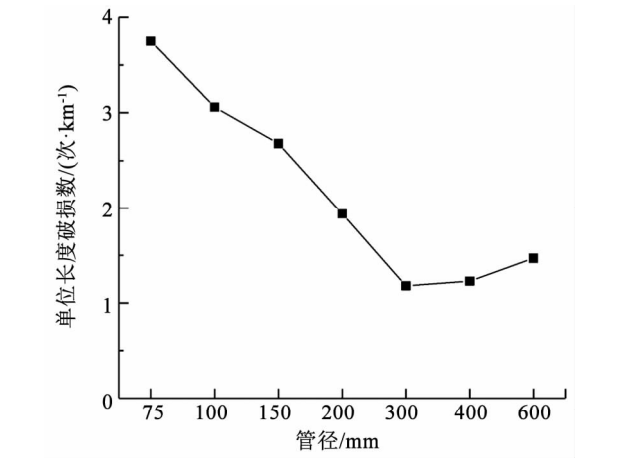


图 1 单位长度破损数随管径变化

Fig. 1 Variation of number of failures per kilometer with pipe diameter

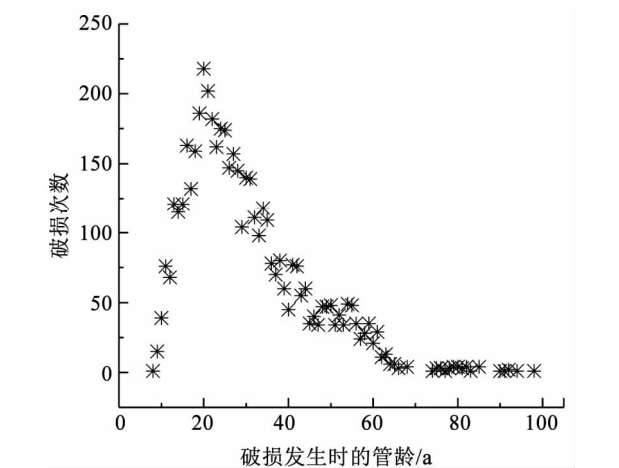


图 2 破损次数随管龄变化

Fig. 2 Variation of number of failures with pipe age

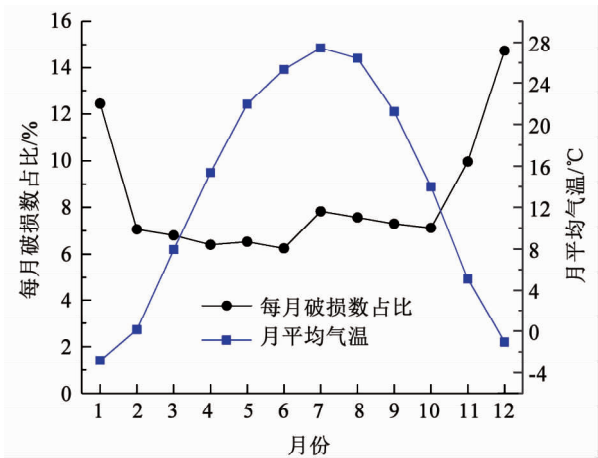


图 3 每月破损数占比及平均气温随月份变化

Fig. 3 Variation of percentage of failures per month and average temperature in different month

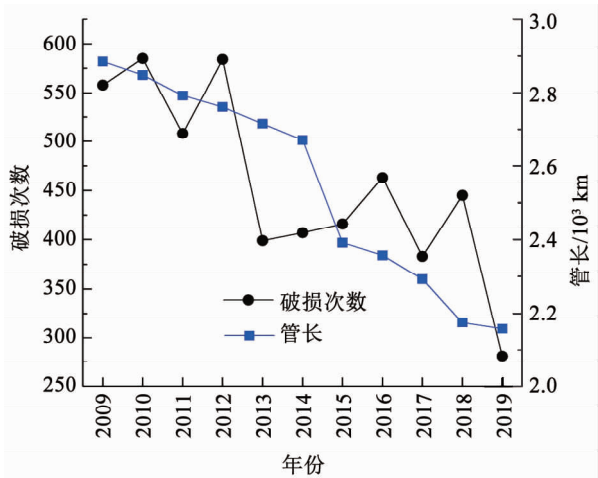


图 4 破损次数及管长随年份变化

Fig. 4 Variation of number of failures and pipe length in different year

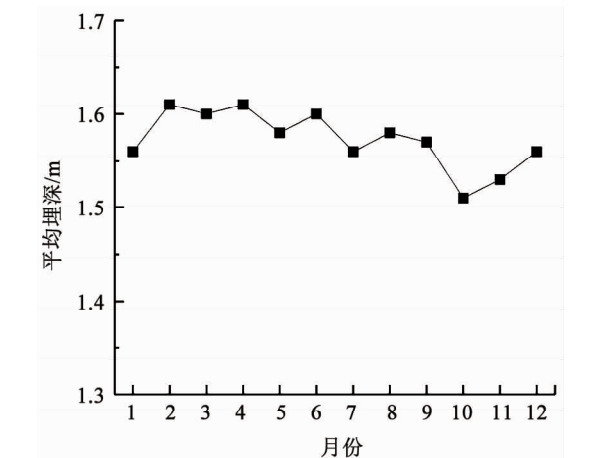


图 5 破损点每月平均埋深

Fig. 5 Average buried depth of failures per month

2 给水管网漏损预测步骤

2.1 模型输入参数选择

在漏损影响因素研究方面,管径(D)、管龄

(A)、管长(L)是模型较为主要的输入参数。Asnaashari 等^[2]采用人工神经网络建立管道漏损率与8个变量(管长、管径、管龄、管材、破损类型、土壤类型、水泥砂浆内衬时间、阴极保护法使用时间)的管道漏损率模型。Sattar 等^[18]采用极限学习机建立年均漏损数与管龄、管长、管径、水泥砂浆保护、阴极保护、历史漏损次数的数学模型。结合图1、2、4中的统计规律,将 D 、 A 、 L 作为漏损影响因素。

对于天气因素(T)的影响,通过 Barton 等^[14]研究天气变化对管道影响的文章结论的综述可知,在气温变化显著的城市,普通铸铁管受天气影响极为明显。同时,结合图3的统计规律可知,该城市漏损次数随着气温降低明显上升,表明普通铸铁管漏损次数与寒冷天气存在明显相关性,故选择天气因素(T)作为漏损影响因素。综上,将 D 、 A 、 L 、 T 确定为模型输入参数。

确定模型输入参数之后,需要通过数据分组得到模型输入输出数据。Xu 等^[12]认为建立模型前需考虑数据分组问题,因为不同的数据分组方法会得到不同模型数据。由图4可知,2008—2019年管道长度逐年下降,故将所有管道先按照年份和管径 D 进行分组(即认为不同管径、管龄的管道受天气因素的影响存在差异),并计算每组内的总管长 L_i 和长度加权管龄 A_i 。

对于某个自然年份中,不同日期的气温是随时间变化的动态数据,且图3所示不同月份的管道漏损数差异明显,因此,在年份、管径分组的基础上,再对1年中的不同月份进行分组,统计不同管径、管龄、月份分组中的天气因素指标、漏损数量。其中按月份分组的日期时间间隔称为时间步长。

由于每日气温数据并不能直接反映天气变化对管道漏损的影响,需要定义天气因素量化指标。采用 Rajani 等^[16]定义的6种天气因素量化指标,具体指标定义和表达式见表2。 $T_{i,j}$ 、 $T_{i,k}$ 、 $T_{i,p}$ 、 $T_{i,q}$ 为时间步长 i 对应的第 j 、 k 、 p 、 q 天平均温度; m 为时间步长 i 包含的天数; θ 为冰冻阈值,连续2 d(含2 d)温度在 θ 以下时才可计算为冰冻指标。

图6展示了确定模型自变量和因变量的处理方法,其中 D 、 A 、 L 、 $I_{T1} \sim I_{T6}$ 、 B 分别为管径、管龄、管长、6个天气指标、时间步长中的破损数。以时间步长 $m = 30$ d为例,在2009-01-01—2009-01-31,根据天气指标定义公式可得该时间步长的6个天气指标 $I_{T1} \sim I_{T6}$ 。

表2 天气因素量化指标定义

Tab. 2 Definitions of quantitative indicators of weather factors

序号	指标名称	表达式
1	平均温度	$I_{T1} = \sum_{j=1}^m T_{i,j}/m$
2	最大上升值	$I_{T2} = \text{Max}\{T_{i,k} - T_{i,j}\}, \forall k > j$
3	最大下降值	$I_{T3} = \text{Max}\{T_{i,q} - T_{i,p}\}, \forall p > q$
4	最大上升率	$I_{T4} = \frac{\text{Max}\{T_{i,k} - T_{i,j}\}}{k - j}, \forall k > j$
5	最大下降率	$I_{T5} = \frac{\text{Max}\{T_{i,q} - T_{i,p}\}}{p - q}, \forall p > q$
6	冰冻指标	$I_{T6} = \sum_{j=1}^m (\theta - T_{i,j}), \forall T_{i,j} < \theta$

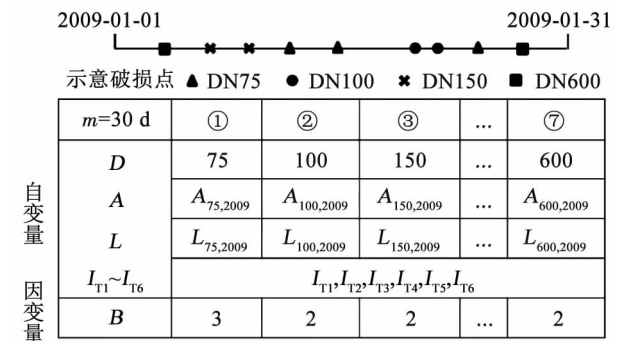


图6 模型输入与输出的定义

Fig. 6 Definitions of model input and output

2.2 模型预测流程

在模型预测流程中,参考 Rajani 等^[16]所采用的分析方法确定天气因素指标的时间步长 m 以及代表性天气指标 I_T^* ,然后采用误差反向传播神经网络(BPNN)和基因表达式编程(GEP)方法建立考虑天气因素的给水管道漏损预测模型。考虑天气因素指标的给水管道漏损预测的具体步骤如下,模型预测流程如图7所示。

1)收集给水管网破损记录数据库、市政管道地理信息数据库和同期气温记录,对数据库的异常数据、缺失数据进行处理。

2)统计不同天气温度下对应的管道破损数,确定天气因素冰冻指标的冰冻阈值 θ 。

3)确定天气因素指标的时间步长 m ,选择代表性天气指标 I_T^* 。

①对30、60、90、120 d等不同的时间步长,按照图6所示的过程计算每个时间步长对应的天气因素指标值,得到不同时间步长对应的建模数据集;

②对于不同时间步长,采用十重交叉验证技术和BPNN方法,分别建立因变量(破损数)与自变量(管径、管龄、管长和单个天气因素指标)的映射关系(预测模型),则对于表2中的6个指标分别建立

6 个预测模型,比较不同预测模型的决定系数,确定最佳时间步长;

③针对最佳时间步长的对应模型,分析天气指标之间相关性,采用枚举法确定余下天气指标的所有组合,对于每种组合模型,共得到 500 个模型决定系数,比较 500 个模型决定系数的均值及变异系数,确定代表性气温表征指标。

4)采用 BPNN 和 GEP 方法建立破损数的隐式和显式预测模型。采用 2008—2018 年的管道破损数据建立模型训练数据集。模型输出(因变量)为管道破损数,模型输入(自变量)包括管径、管龄、管长、代表性天气因素指标。

5)以未来一年(2018 年 8 月—2019 年 7 月)的管道破损数据作为预测数据,验证预测模型的准确性。

采用决定系数(R^2)作为模型预测性能评价指标,定义见式(1)。 R^2 反映因变量的全部变异性能通过回归关系被自变量解释的比例, R^2 越接近 1,说明模型性能越好。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{1}$$

式中 y_i 、 $\hat{y}_i$ 、 $\bar{y}$ 分别为预测数据集中第 i 个预测值、实际值、所有预测值的均值。

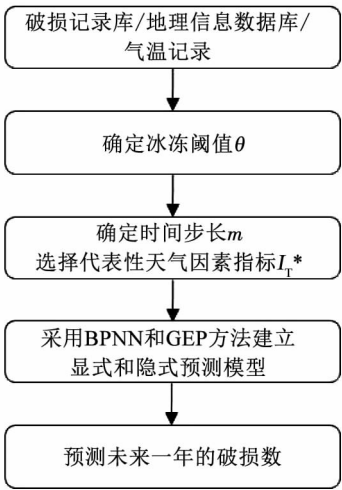


图 7 模型预测流程

Fig. 7 Flow chart of prediction model

2.3 构建预测模型的方法

2.3.1 误差反向传播神经网络

误差反向传播神经网络(BPNN)是一种前向多层反向传播学习算法。在前向传递中,输入信号从输入层经隐含层逐层处理,直至输出层。每一层的神经元状态只影响下一层神经元状态。如果输出层

得不到期望输出,则转入反向传播,根据预测误差调整网络权值和阈值,从而使 BPNN 预测输出不断逼近期望输出。图 8 为基于 BPNN 构建因变量 Y 与自变量 $X = \{X_1, X_2, \dots, X_n\}$ 函数映射关系的原理示意, ω_{ij} 和 ω_{jk} 为 BPNN 的权值。可以看出,将 BPNN 看作是一个非线性函数,输入值和输出值分别为函数的自变量和因变量。当输入节点数为 n 、输出节点数为 N 时,BPNN 就表达了从 n 个自变量到 N 个因变量的函数映射关系^[19]。BPNN 预测流程分为模型构建、模型训练、模型预测 3 大步骤。

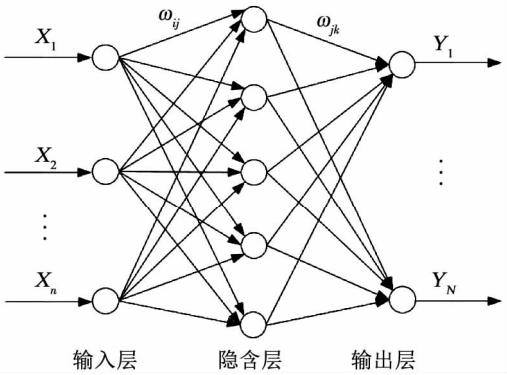


图 8 BPNN 拓扑结构图

Fig. 8 BPNN topology diagram

2.3.2 基因表达式编程

基因表达式编程(GEP)是 Ferreira^[20]于 2001 年在传统的遗传算法(genetic algorithm, GA)和遗传编程(genetic programming, GP)的基础上提出的一种新型演化算法,其主要特点在于可以建立因变量与自变量之间的显式函数关系式。在 GP 中,个体由不同大小和形状的表达式树表示,但其遗传操作是直接作用于 GP 的树状编码上,随着进化的进行就会出现代码膨胀问题,进化效率越来越低,导致不利于进行遗传操作。GEP 的基因具有两种表现形式,即基因型和表现型。任何遗传操作算子的操作都是在基因型上进行(GA 优势),而表现型则是基因型的解码,根据求解问题的性质将固定长度的线性结构个体解码成一棵表达式树(GP 优势),从而实现了利用简单编码来解决复杂问题。图 9 举例展示 GEP 和 GP 进行个体“变异”遗传操作时两者的差异。GEP 只需在基因第 4 个位置上进行变异操作,然后解码成表达式树即得新的子代。GP 进行变异操作时需要删除表达式树突变点“b”及其以下的树分支,然后用表达式树“a”代替,才得到新子代。

GEP 算法的基本过程和 GA 与 GP 相似,首先随机生成含有一定数量个体的染色体(初始种群),然后染色体解码成表达式树,并评价这些染色体的适

应度,接着通过适应度函数值等指标对解析获得的表达式树进行评价,判断其是否满足进化约束条件。若未满足,则采用复制、变异、重组等方式进行遗传进化处理,从而得到新一代染色体。新一代染色体进入下一轮的生存迭代过程,评价适应度、被选择、经历遗传操作等。这一过程反复进行下去,直到迭代终止条件满足为止^[21]。

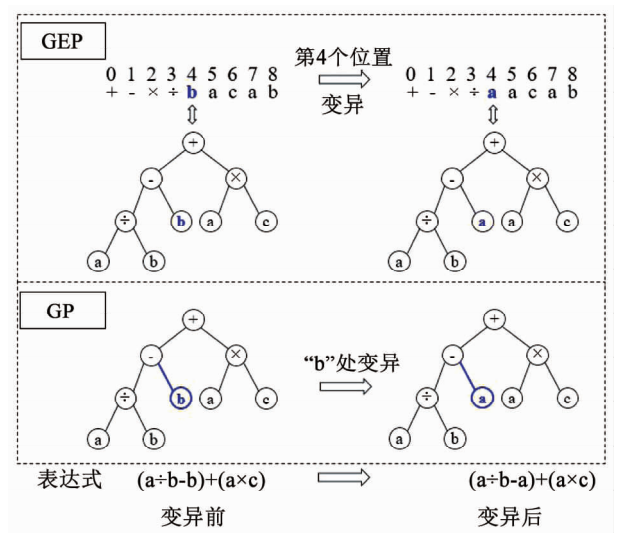


图9 GEP和GP的遗传操作

Fig. 9 Genetic manipulation of GEP and GP

3 实例分析

3.1 冰冻指标阈值

根据2.2节步骤建立案例城市给水管道漏损预测模型。对基础数据进行处理后的统计分析,结果见图1~5。第2)步确定冰冻指标阈值 θ 时,图3中2月份和11月份的气温和破损数变化说明 θ 取值应该在0~6℃。基于11年气温数据,统计不同气温对应天数以及发生的破损数,计算得到不同气温对应日平均破损数,结果如图10。可以看出,0~6℃气温从3℃变化为4℃时,对应的日平均破损数存在明显下降。综上,将冰冻阈值确定为 $\theta=3℃$ 。

3.2 时间步长及代表性天气指标

分别采用 $m=\{30,60,90,120\}$ d的时间步长,对于每个时间步长,采用BPNN方法分别建立自变量(管径、管龄、管长、单个天气指标)与因变量(管道破损数)的预测模型,即每个时间步长建立6种预测组合。为避免BPNN建模结果存在偶然性,每次BPNN模型构建时,采用十重交叉验证方法组建模型训练数据集,则可得到10组不同的训练数据集;对每组训练数据,独立运行50次BPNN建模,则每种预测组合最终得到500个隐式预测模型。500个预测模型的拟合结果决定系数的均值(μ)和变异系

数(c_v)见表3。其中, μ 值越大表明模型预测精度越高, c_v 越小表明模型预测结果越稳定。

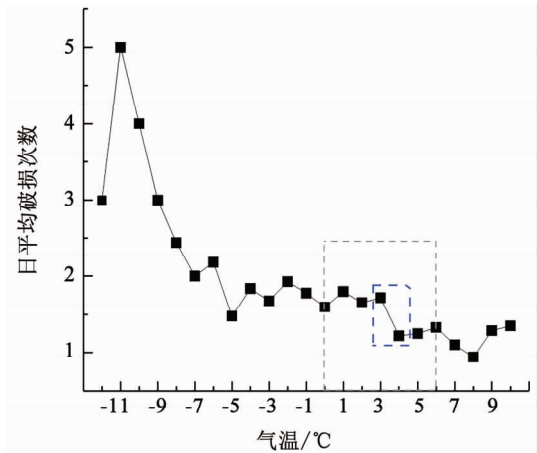


图10 日平均破损次数随气温变化

Fig. 10 Variation of average daily failures with temperature

表3 模型决定系数均值和变异系数

Tab. 3 Mean value and variation coefficient of model determination coefficient

m	μ/c_v			
	30 d	60 d	90 d	120 d
I_{T1}	0.68/13.72	0.77/9.51	0.82 /14.24	0.80/17.99
I_{T2}	0.43/16.12	0.63/13.78	0.67/13.27	0.76 /12.64
I_{T3}	0.56/15.88	0.61/18.07	0.69/14.74	0.74 /14.71
I_{T4}	0.57/14.01	0.59/41.80	0.63/33.38	0.72 /20.94
I_{T5}	0.58/13.44	0.64/19.52	0.74 /38.19	0.68/35.91
I_{T6}	0.68/8.88	0.76/12.55	0.84/9.61	0.85 /7.23

由表3可知,随着时间步长(m)的增加,预测模型决定系数整体呈现上升趋势,即模型拟合精度提升。但对于部分天气因素指标,该上升趋势并不连续。在 m 由90 d变为120 d时,天气指标 I_{T1} (平均温度)、 I_{T5} (最大下降率)对应的模型精度则呈现下降趋势。由于图3所示管道漏损事件在11月一次年1月有明显增加,且根据表3中不同模型决策系数均值和变异系数的结果,选取90 d作为案例城市管道预测模型的时间步长。

基于 $m=90$ d,计算案例城市2008—2018年的6个天气因素指标值,这些天因素指标的Pearson相关系数计算结果见表4。表中平均温度(I_{T1})与冰冻指标(I_{T6})的相关系数为-0.89,两指标存在显著的负相关性。因此,在进行预测模型对天气指标的敏感度分析时,剔除平均温度指标,对剩余的5个天气因素指标进行分析。

表 4 天气指标的相关系数

Tab. 4 Correlation coefficients of weather variables						
相关系数	I_{T1}	I_{T2}	I_{T3}	I_{T4}	I_{T5}	I_{T6}
I_{T1}	1.00	-0.04	-0.02	0.15	0.21	-0.89
I_{T2}		1.00	-0.44	0.15	0.37	-0.16
I_{T3}			1.00	0.46	0.15	0.16
I_{T4}				1.00	0.35	-0.07
I_{T5}					1.00	-0.17
I_{T6}						1.00

采用枚举法确定 5 个天气因素指标的所有组合,共得到 31 种组合方式($C_5^1 + C_5^2 + C_5^3 + C_5^4 + C_5^5 =$

31)。对每种天气指标组合,以 $\{D,A,L,$ 天气指标组合 $\}$ 为自变量,使用十重交叉验证和独立运行 50 次的方法,基于 BPNN 分别构建每个组合方式的预测模型。对于每种天气指标组合,预测模型决定系数的均值和变异系数如图 11 所示。可以看出,第 5、12 种天气组合的模型预测精度较高,决定系数平均值均为 0.86,决定系数的变异系数分别为 5.38% 和 7.33%。因为两种组合模型精度较接近,优先选择自变量较少的组合(第 5 种组合,天气指标为 I_{T6})。综上,选择冰冻指标为代表性气温表征指标。

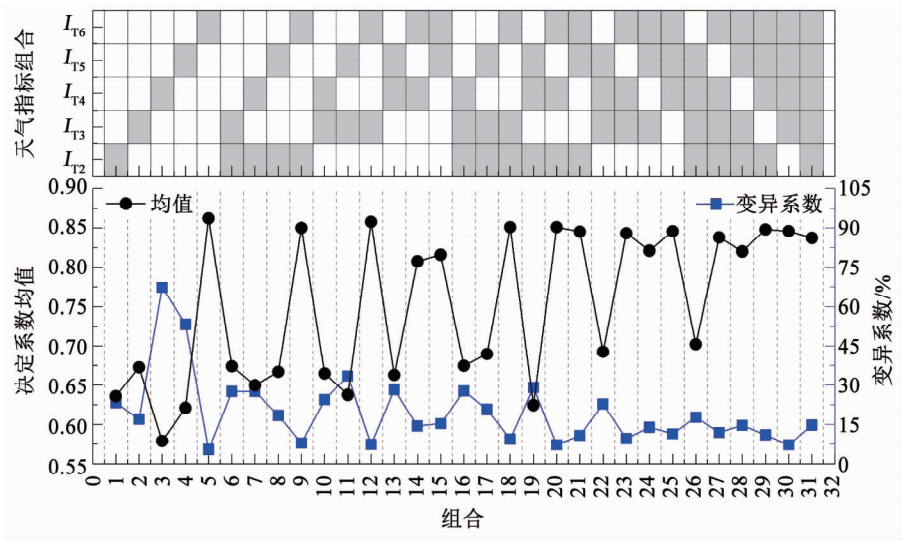


图 11 组合模型的决定系数均值及变异系数

Fig. 11 Mean value and variation coefficient of determination coefficient of different combination models

3.3 构建预测模型

通过以上分析,最终确定模型时间步长为 90 d,模型输入变量(自变量)为管径(D)、管龄(A)、管长(L)、冰冻指标(I_F),输出变量(因变量)为时间步长内的管道破损数(B)。将案例城市 2008-08-01—2018-07-31 的漏损记录数据按照管径、管龄、时间步长进行分组,共得到 308 组数据,其中 280 组数据为模型训练数据,其余 28 组数据为模型测试数据。

建立因变量与自变量之间的隐式函数模型时,

采用 3 层结构的 BPNN,其中输入层和输出层节点数分别为 4 和 1,隐含层节点数 N 由经验公式 $N = 2n + a$ 得出, n 为输入层节点数, a 为 1~10 的常数^[13, 22]。通过试凑法确定 $a = 7$ 时模型预测误差较小,此时对应隐含层节点数 $N = 15$,即最后确定网络模型结构为 4-15-1。数据归一化和 BPNN 模型的构建采用 MATLAB 机器学习工具箱的内置函数。采用 GEP 建立因变量与自变量之间的显式函数模型时,参数设置见表 5。

表 5 GEP 建模参数设置
Tab. 5 Parameters for GEP modeling

参数名称	进化代数	染色体个数	基因个数	头部长度	连接函数	适应度函数	函数集	变异率	重组率
参数设置	2 000	30,40,50	1,2,3	7,8,9,10	+,×	RMSE	+, -, ×, ÷, exp, ln, Inv, sqrt, x^2	0.005~0.05	0.1,0.2,0.3

采用 BPNN 建立隐式预测模型时,采用十重交叉验证和 50 次独立运行,共生成 500 个模型,选取模型测试集预测结果决定系数最优的模型作为选定的模型。GEP 在建模过程中独立运行 5 次,共生成 774 个模型表达式,根据表达式结构和模型测试集预测结果决定系数挑选出 5 个模型,表达式如表 6

所示。 R_{TR}^2 和 R_{TE}^2 分别为模型训练集和测试集决定系数。可以看出,模型②④决定系数相对较低,模型⑤虽然决定系数最好,但模型表达式太过复杂,模型①③表达式结构较相近,但模型①预测精度较高。此外,模型①表达式中漏损数量 B 与管径 D 为反比关系, B 与管道长度 L 、冰冻指标 I_F 为正比关系,这

些规律与图 1、3、4 中的结论一致。故最终选择模型①的表达式为显式预测模型。

表 6 GEP 模型表达式				
Tab. 6 Equations obtained by GEP				
序号	GEP 模型表达式	R^2_{TR}	R^2_{TE}	
①	$B = \frac{3.1[2.8L + I_F(1 + A^{-1}) - 3A]}{2A + D - 21.5}$	0.83	0.91	
②	$B = \frac{4.8(2L - A + 0.8I_F - 0.2D)}{83.7 - \sqrt{L} + D}$	0.85	0.74	
③	$B = \frac{(L + 10.8)[8 - (0.1I_F + 0.8)^{-1}]}{A + D - 1.7\sqrt{I_F}}$	0.83	0.88	
④	$B = \frac{A^2 + I_F + 692}{32.53A^2} \left(L - \frac{9.924 + D}{3.18} \right)$	0.74	0.79	
⑤	$B = \frac{0.3\ln[L(L + D) - A^2(I_F + D)]}{(4.7 + D + A)(2L + I_F + A + 18.3)^{-1}}$	0.86	0.93	

3.4 模型预测结果分析

采用 BPNN 和 GEP 方法建立的考虑天气影响指标的管道漏损预测隐式 (BP-T) 和显式 (GEP-T) 模型, 预测案例管网在 2018 年 8 月—2019 年 7 月的管道破损数, 结果见图 12。此外, 图 12 还给出了采用 BPNN 和 GEP 方法建立的未考虑天气指标的隐式 (BP) 和显式 (GEP) 模型预测结果的对比。其中, 预测时段的管道破损记录按照管龄、管径、时间步长进行分组, 共得到 28 组实测数据。如图 12 所示, 每个管径包含 4 组数据, 分别对应时间步长 8—10 月 (T1), 11 月—次年 1 月 (T2), 2—4 月 (T3), 5—7 月 (T4)。

由图 12 可知, 未考虑天气因素指标的显式和隐

式模型预测决定系数分别为 0.65 和 0.60, 考虑天气因素指标的显式和隐式模型预测结果的决定系数分别为 0.78 和 0.88。考虑天气因素后, 显式和隐式模型的预测精度提升率分别为 13% 和 28%。因此, 考虑天气因素可有效提升漏损预测模型的整体精度。由图 12 中“实际记录值”可知, 天气因素对 DN75 ~ DN150 分组的管道漏损数存在显著影响, 即在 T2 分析步 (11 月—次年 1 月) 中存在管道漏损数量上升, 本研究建立的 BP-T 模型可以提高预测模型在这些分析步中的预测精度。而对于 DN200 ~ DN600 分组的分析步中, 并未表现出天气因素对 T2 分析步漏损数的影响, 在这些分组中, 考虑天气因素的预测模型 BP-T 和 GEP-T 并未表现出较高的预测精度。因此, 天气因素指标对预测模型精度的提升, 适用于实际漏损记录中显著存在天气影响的分组数据。由于实际管网中不同分组类型管道漏损的主要影响因素存在差异, 不可简单地预期考虑天气因素指标可以提高所有类型管道漏损预测的精度, 也不应仅采用天气因素指标建立预测模型。

图 12 在考虑天气因素指标的模型中, BPNN 模型的预测精度高于 GEP 模型, 而在未考虑天气因素指标的模型中, GEP 模型的预测精度高于 BPNN 模型。这说明不同建模方法的适用性随着模型参数数据而变化, 在构建给水管道漏损预测模型时, 有必要采用不同的建模方法, 以获得对于案例管网数据较为客观全面的认知。因此, 本研究同时采用显式和隐式方法构建模型的思路是可行的。

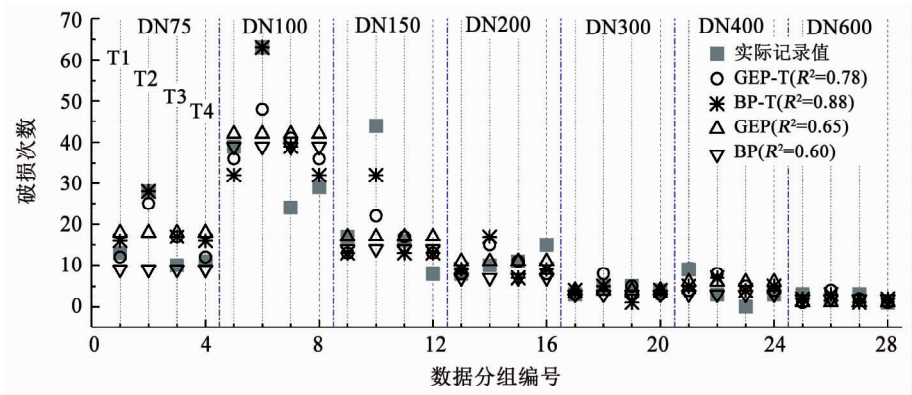


图 12 模型预测值与实际值比较

Fig. 12 Comparison between model prediction values and observation values

4 结 论

基于北方某城市给水管网破损事件数据和气温记录, 分析了不同的天气因素量化指标与管道破损事件的相关性, 采用反向传播人工神经网络

(BPNN) 和基因表达式编程 (GEP) 算法, 建立了考虑天气因素的给水管道漏损预测模型。所建立的模型可预测未来 1 年不同时段管道破损数, 为管道更新改造、运营维护等管道漏损事件发生前的干预工作提供指导。主要结论如下:

1)通过模型结果判断最佳时间步长为 90 d,结合相关性分析及模型预测结果确定冰冻指标为天气因素代表性指标,考虑冰冻指标后的显式和隐式模型预测精度提升率分别为 13% 和 28%,说明考虑天气因素可明显提高漏损预测模型的拟合精度。

2)天气因素指标对预测模型精度的提升,适用于实际漏损记录中显著存在天气影响的分组数据,由于实际管网中不同分组类型管道漏损的主要影响因素存在差异,不可简单地预期考虑天气因素指标可以提高所有类型管道漏损预测的精度,也不应仅采用天气因素指标建立预测模型。

3)在考虑天气因素指标的模型中,BPNN 模型的预测精度高于 GEP 模型,而在未考虑天气因素指标的模型中,GEP 模型的预测精度高于 BPNN 模型,说明不同建模方法的适用性随着模型参数数据而变化,采用不同的建模方法将有利于获得对管网数据较为客观全面的认知。

参考文献

- [1] SATTAR A M A, GHARABAGHI B, MCBEAN E A. Prediction of timing of watermain failure using gene expression models[J]. *Water Resources Management*, 2016, 30 (5): 1635. DOI: 10.1007/s11269-016-1241-x
- [2] ASNAASHARI A, MCBEAN E A, GHARABAGHI B, et al. Forecasting watermain failure using artificial neural network modelling [J]. *Canadian Water Resources Journal*, 2013, 38(1): 24. DOI: 10.1080/07011784.2013.774153
- [3] DAVIS P, SILVA D D, MARLOW D, et al. Failure prediction and optimal scheduling of replacements in asbestos cement water pipes [J]. *Journal of Water Supply: Research and Technology-Aqua*, 2008, 57(4): 239. DOI: 10.2166/aqua.2008.035
- [4] XU Q, CHEN Q, LI W. Application of genetic programming to modeling pipe failures in water distribution systems[J]. *Journal of Hydroinformatics*, 2011, 13(3): 419. DOI: 10.2166/hydro.2010.189
- [5] LIN P, YUAN X. A two-time-scale point process model of water main failures for infrastructure asset management [J]. *Water Research*, 2019, 150: 296. DOI: 10.1016/j.watres.2018.11.066
- [6] KLEINER Y, RAJANI B. I-WARP: individual water main renewal planner[M]. Boca Raton: CRC Press-Taylor & Francis Group, 2010:639
- [7] ANDREOU S A, MARKS D H, CLARK R M. A new methodology for modelling break failure patterns in deteriorating water distribution systems: theory[J]. *Advances in Water Resources*, 1987, 10(1): 2. DOI: 10.1016/0309-1708(87)90002-9
- [8] 赵洪宾, 陈兵, 伍悦滨. 给水管网漏失预测模型的研究[J]. *给水排水*, 2001, 27(10): 94
ZHAO Hongbin, CHEN Bing, WU Yuebin. Research on leakage prediction model of water supply network[J]. *Water & Wastewater Engineering*, 2001, 27 (10): 94. DOI: 10.13789/j.cnki.wwe1964.2001.10.033
- [9] 张宏伟, 牛志广, 陈超. 供水管网漏损预测模型研究[J]. *中国给水排水*, 2001, 17(6): 7
- ZHANG Hongwei, NIU Zhiguang, CHEN Chao. Study on the prediction model for water supply net leakage[J]. *China Water & Wastewater*, 2001, 17(6): 7. DOI: 10.3321/j.issn:1000-4602.2001.06.003
- [10] 邵圆媛. 嵌套 BP/GMS 神经网络模型在供水管网漏损预测中的研究[D]. 重庆: 重庆大学, 2018
SHAO Yuanyuan. Research on nested BP/GMS neural network models in water supply network leakage prediction[D]. Chongqing: Chongqing University, 2018
- [11] 徐强, 陈求稳, 李伟峰, 等. 基于遗传编程的供水管网漏失模型[J]. *中国给水排水*, 2011, 27(1): 45, 51
XU Qiang, CHEN Qiuwen, LI Weifeng, et al. Development of pipe leakage model for water distribution system using genetic programming[J]. *China Water & Wastewater*, 2011, 27(1): 45, 51. DOI: 10.19853/j.zgjsps.1000-4602.2011.01.012
- [12] XU Q, CHEN Q, LI W, et al. Pipe break prediction based on evolutionary data-driven methods with brief recorded data [J]. *Reliability Engineering & System Safety*, 2011, 96(8): 942. DOI: 10.1016/j.res.2011.03.010
- [13] 阎立华, 马婷婷, 田昊平. BP 神经网络对管网安全使用期的预测[J]. *沈阳建筑大学学报(自然科学版)*, 2008, 24(5): 845
YAN Lihua, MA Tingting, TIAN Haoping. Research on prediction of safe using dates of water supplying pipes based on BP neural network [J]. *Journal of Shenyang Jianzhu University (Natural Science)*, 2008, 24(5): 845
- [14] BARTON N A, FAREWELL T S, HALLETT S H, et al. Improving pipe failure predictions: factors affecting pipe failure in drinking water networks[J]. *Water Research*, 2019, 164: 114926. DOI: 10.1016/j.watres.2019.114926
- [15] 邓晓婷. 城市供水管网漏损因素分析及控制[D]. 太原: 太原理工大学, 2012
DENG Xiaoting. Urban water supply pipe network leakage factor analysis and control [D]. Taiyuan: Taiyuan University of Technology, 2012
- [16] RAJANI B, KLEINER Y, SINK J. Exploration of the relationship between water main failures and temperature covariates[J]. *Urban Water Journal*, 2012, 9(2): 67. DOI: 10.1080/1573062X.2011.630093
- [17] KAKOUDAKIS K, FARMANI R, BUTLER D. Pipeline failure prediction in water distribution networks using weather conditions as explanatory factors[J]. *Journal of Hydroinformatics*, 2018, 20(5): 1191. DOI: 10.2166/hydro.2018.152
- [18] SATTAR A M A, ERTUGRUL O F, GHARABAGHI B, et al. Extreme learning machine model for water network management[J]. *Neural Computing and Applications*, 2019, 31(1): 157. DOI: 10.1007/s00521-017-2987-7
- [19] 王小川. MATLAB 神经网络 43 个案例分析[M]. 北京: 北京航空航天大学出版社, 2013
- [20] FERREIRA C. Gene expression programming: a new adaptive algorithm for solving problems[J]. *Complex Systems*, 2001, 13(2): 87
- [21] 元昌安. 基因表达式编程算法原理与应用[M]. 北京: 科学出版社, 2010
- [22] HECHT-NIELSEN R. Theory of the backpropagation neural network [M]. New York: Academic Press, 1989